



## LINGVISTIK KORPUSLARDA ONTOLOGIYADAN FOYDALANISH MASALALARI

**Mamarajabov Abduvali Yuldosh o'g'li**

TerDU O'zbek tili va adabiyoti kafedrası o'qituvchisi

**Annotatsiya.** Ushbu maqolada lingvistik korpuslarda ontologiyadan foydalanishning nazariy va amaliy jihatlari tahlil qilinadi. Ontologiya korpus ma'lumotlarini semantik jihatdan tizimlashtirish, so'zlar va tushunchalar o'rtasidagi munosabatlarni aniqlashda muhim vosita sifatida ko'riladi. Shuningdek, korpus lingvistikasida avtomatik tahlil jarayonlarini takomillashtirishda ontologik modellarning o'rni yoritiladi. Tadqiqotda mavjud yondashuvlar va muammolar ham muhokama qilinadi.

**Kalit so'zlar:** korpus lingvistikasi, ontologiya, semantik tahlil, ma'lumotlar bazasi, bilimlar tizimi, avtomatik tahlil, lingvistik model, NLP, leksik munosabat, formal tasvirlash.

### Kirish

Zamonaviy tilshunoslikda lingvistik korpuslar til birliklarini real kontekstda o'rganishning eng muhim vositalaridan biri hisoblanadi. Korpus ma'lumotlarini chuqur tahlil qilish esa faqat statistik yondashuv bilan cheklanib qolmay, semantik va konseptual modellashtirishni ham talab etadi. Shu nuqtayi nazardan ontologiya korpus lingvistikasida muhim metodologik asos bo'lib xizmat qiladi. Ontologiya yordamida til birliklari o'rtasidagi mantiqiy va semantik aloqalar tizimlashtiriladi, bu esa avtomatik tahlil, mashina tarjimai va sun'iy intellekt tizimlarining samaradorligini oshiradi. Mazkur maqolada lingvistik korpuslarda ontologiyadan foydalanishning asosiy yo'nalishlari, imkoniyatlari va mavjud muammolari tahlil qilinadi.

### Asosiy qism

Zamonaviy tilshunoslikda lingvistik korpuslar til birliklarini tabiiy kommunikativ muhitda o'rganish imkonini beruvchi eng muhim empirik baza sifatida e'tirof etiladi. Korpus lingvistikasi rivojlanishi bilan birga, til ma'lumotlarini faqat statistik yoki strukturaviy tahlil qilish yetarli bo'lmay qoldi. Bugungi kunda semantik, konseptual va bilimlar bazasiga asoslangan yondashuvlar dolzarblik kasb etmoqda. Ana shunday yondashuvlardan biri – ontologiyaga asoslangan lingvistik modellashtirishdir.

Ontologiya, umumiy ma'noda, ma'lum bir soha bilimlarini formal va ierarxik tizimda ifodalovchi konseptual model sifatida tushuniladi. Lingvistik korpuslar bilan integratsiyada esa ontologiya til birliklari o'rtasidagi semantik munosabatlarni aniq



belgilash, ma'no qatlamlarini tizimlashtirish va avtomatik tahlil jarayonlarini takomillashtirish vositasi sifatida namoyon bo'ladi [4, 12].

Lingvistik korpuslar odatda katta hajmdagi matnlar to'plamidan iborat bo'lib, ular morfologik, sintaktik va leksik birliklarni o'z ichiga oladi. Biroq bunday korpuslarda ma'no qatlamini chuqur tushunish uchun qo'shimcha semantik struktura zarur bo'ladi. Ontologiya aynan shu bo'shliqni to'ldiradi.

C. Chiarcos tomonidan ishlab chiqilgan lingvistik annotatsiyalar ontologiyasi til birliklarini formal kategoriyalarga ajratish va ularning o'zaro munosabatlarini ko'rsatish imkonini beradi. Ushbu yondashuvga ko'ra, har bir til elementi (so'z, morfema, sintaktik birlik) ma'lum bir konseptual tugun sifatida qaraladi va ular orasidagi munosabatlar "is-a", "part-of", "related-to" kabi semantik bog'lanishlar orqali ifodalanadi [4, 15]. Bu yondashuv korpus lingvistikasida nafaqat tavsifiy, balki tushuntiruvchi (explanatory) model yaratish imkonini beradi.

Ontologik modellashtirish lingvistik ma'lumotlarni formal tizimga keltirishni nazarda tutadi. Bu jarayon uch asosiy bosqichdan iborat:

1. konseptlarni aniqlash
2. konseptlar o'rtasidagi munosabatlarni belgilash
3. bilimlar bazasini formal ko'rinishda ifodalash

C. M. Keetning tadqiqotlariga ko'ra, ontologiya yaratish jarayoni faqat texnik emas, balki epistemologik jarayon ham hisoblanadi, chunki u til va bilim o'rtasidagi chegarani aniqlaydi [5, 23]. Lingvistik korpuslarda bu jarayon avtomatik yoki yarim avtomatik usullar yordamida amalga oshiriladi. Masalan, mashina o'rganishi algoritmlari asosida matndan terminlar ajratib olinadi va ular ontologik tugunlarga aylantiriladi.

Ontologiya lingvistik korpuslarda semantik tahlilning asosiy vositasi sifatida xizmat qiladi. Semantik tahlil jarayonida so'zlarning faqat shakli emas, balki ularning kontekstual ma'nosi ham hisobga olinadi. Ontologiya esa bu ma'nolarni tizimlashtiradi.

Abjalova M.ning tadqiqotlarida ta'kidlanishicha, korpus lingvistikasida semantik qatlamni aniqlash uchun ontologik yondashuv eng samarali usullardan biri hisoblanadi, chunki u til birliklarini faqat lingvistik emas, balki konseptual obyekt sifatida ham ko'rib chiqadi [2, 18]. Natijada, korpus ma'lumotlari asosida yaratilgan tizimlar nafaqat so'zlarni qidirish, balki ularning ma'nodoshligi, antonimligi va konseptual aloqalarini ham aniqlay oladi.

O'zbek tilshunosligida ham korpus lingvistikasi va ontologiya integratsiyasi bo'yicha tadqiqotlar olib borilmoqda. Abduraxmonova N. tomonidan ishlab chiqilgan elektron



korpus modeli o'zbek tilining formal-funksional xususiyatlarini hisobga olgan holda qurilgan bo'lib, unda ontologik yondashuv asosiy metodologik vosita sifatida qo'llanilgan [1, 7].

Ushbu modelda har bir til birligi korpus ichida mustaqil obyekt sifatida emas, balki ma'lum bir konseptual tizimning elementi sifatida qaraladi. Bu esa o'zbek tilining raqamli resurslarini yanada boyitish va avtomatik tahlil imkoniyatlarini kengaytiradi.

### **Ontologiyaning amaliy qo'llanilishi**

Ontologiya lingvistik korpuslarda quyidagi amaliy yo'nalishlarda qo'llaniladi:

- mashina tarjima tizimlari
- axborot qidiruv tizimlari
- sun'iy intellekt asosidagi NLP modellar
- avtomatik annotatsiya tizimlari

Masalan, semantik qidiruv tizimlarida foydalanuvchi so'roviga faqat kalit so'z emas, balki uning ma'nodosh va konseptual bog'liq birliklari ham hisobga olinadi. Bu esa qidiruv aniqligini sezilarli darajada oshiradi. Keetning ta'kidlashicha, ontologik modellar asosida qurilgan tizimlar an'anaviy leksik qidiruv tizimlariga qaraganda 30–40% yuqori aniqlik beradi [5, 31].

Ontologiyadan foydalanishning bir qator muammolari ham mavjud. Ular quyidagilardan iborat:

- lingvistik noaniqlik (ambiguity)
- ko'p ma'nolilikni formal ifodalash qiyinligi
- avtomatik ontologiya yaratishdagi xatoliklar
- resurslarning yetishmasligi

Ayniqsa, tabiiy tilning dinamikligi ontologik modellashtirishni murakkablashtiradi. Har bir til doimiy rivojlanishda bo'lgani uchun ontologiya ham muntazam yangilanib borishi talab etiladi.

Lingvistik korpuslarda ontologiyadan foydalanish jarayoni dastlab oddiy semantik tasniflash darajasida boshlangan bo'lsa, bugungi kunda u to'liq bilimlar muhandisligi (knowledge engineering) yo'nalishiga aylangan. Dastlabki korpuslar asosan statistik tahlilga asoslangan bo'lib, so'zlarning chastotasi, kollokatsiyalari va sintaktik strukturasi o'rganilgan. Biroq bu yondashuv tilning chuqur semantik qatlamini ochib bera olmagan. Ontologik modelning kirib kelishi bilan korpuslar endi faqat "matnlar ombori" emas, balki "bilimlar tizimi" sifatida qaralmoqda. Bu o'zgarish korpus lingvistikasining paradigmatic siljishini ifodalaydi: tavsifiy (descriptive) yondashuvdan konseptual va formal (formal-semantic) yondashuvga o'tish amalga oshdi [4, 22]. Bu jarayonda lingvistik birliklar endi alohida so'z yoki struktura sifatida



emas, balki ma'lum bir bilim domenining elementlari sifatida qaraladi. Masalan, "o'qituvchi", "ta'lim", "metodika" kabi birliklar faqat leksik birlik emas, balki "education ontology"ning o'zaro bog'langan tugunlari sifatida modellashtiriladi. Zamonaviy NLP tizimlarida korpus va ontologiya integratsiyasi uch darajada amalga oshiriladi:

1. Leksik daraja integratsiya. Bu bosqichda korpusdagi so'zlar ontologik lug'atlarga (lexical ontology) moslashtiriladi. Har bir so'zga semantik atributlar biriktiriladi. Bu jarayon ko'pincha WordNet tipidagi resurslar asosida amalga oshiriladi.
2. Sintaktik-semantik daraja integratsiya. Bu darajada gap tuzilmalari ontologik munosabatlar bilan bog'lanadi. Masalan, "A B ni o'rgatadi" strukturasi "o'rgatish" fe'li "education\_event" konseptiga ulanadi va agent-obyekt munosabatlari aniqlanadi.
3. Konseptual daraja integratsiya. Eng yuqori bosqichda butun korpus bilimlar grafigi sifatida qayta quriladi. Bu yerda ontologiya korpusni grafik ma'lumotlar bazasiga aylantiradi.

Keetning tadqiqotlarida ta'kidlanishicha, bunday integratsion model NLP tizimlarining semantik aniqligini sezilarli darajada oshiradi va bilimlarni qayta ishlash samaradorligini kuchaytiradi [5, 41].

Sun'iy intellekt tizimlarining rivojlanishi lingvistik korpuslar bilan bevosita bog'liqdir. Ayniqsa, ontologik korpuslar mashina o'rganishi (machine learning) va chuqur o'rganish (deep learning) modellarining semantik qatlamini boyitishda muhim rol o'ynaydi. An'anaviy neyron tarmoqlar asosan statistik bog'lanishlarga asoslanadi, ya'ni ular so'zlarning ma'nosini emas, balki ularning birga uchrashish ehtimolini o'rganadi. Ontologiya esa bu jarayonga "ma'no strukturasi" qo'shadi. Natijada model faqat "nimadir bilan birga kelgan so'z"ni emas, balki "nimani anglatishini" ham tushuna boshlaydi. Masalan, "shifokor" va "kasalxona" o'rtasidagi bog'lanish faqat statistik emas, balki "medical ontology"ga asoslangan konseptual munosabat sifatida talqin qilinadi.

O'zbek tilshunosligida korpus lingvistikasini rivojlantirish so'nggi yillarda jadallashgan. Abduraxmonova N. va boshqa tadqiqotchilar tomonidan yaratilgan elektron korpuslar bu sohadagi ilk tizimli qadamlar hisoblanadi [1, 9].

Biroq ontologik modellashtirish darajasida hali bir qator muammolar mavjud:

- standartlashtirilgan ontologiya yo'qligi
- milliy til resurslarining cheklanganligi
- annotatsiya tizimlarining to'liq avtomatlashtirilmaganligi

Shu bilan birga, istiqbolli yo'nalishlar ham aniq ko'rinmoqda. Jumladan:

- o'zbek tilining milliy ontologiyasini yaratish



- korpusni semantik grafik bazaga aylantirish
- NLP tizimlarini lokal til xususiyatlariga moslashtirish

Bu yo‘nalishlar amalga oshirilsa, o‘zbek tilining raqamli ekotizimi sezilarli darajada rivojlanadi.

Ontologik yondashuv nafaqat texnik, balki metodologik inqilob ham hisoblanadi. U tilshunoslikni formal fanlar bilan yaqinlashtiradi. Natijada til faqat ijtimoiy hodisa emas, balki formal tizim sifatida ham o‘rganila boshlaydi. Chiarcos tomonidan ishlab chiqilgan lingvistik annotatsiya ontologiyasi shuni ko‘rsatadiki, til birliklarini universal kategoriyalar asosida tavsiflash mumkin. Bu esa turli tillar uchun yagona semantik model yaratish imkonini beradi [4, 27].

### **Xulosa**

Lingvistik korpuslarda ontologiyadan foydalanish til ma’lumotlarini faqat statistik emas, balki semantik va konseptual darajada tahlil qilish imkonini beradi. Ontologiya korpus tarkibidagi til birliklari o‘rtasidagi mantiqiy va ma’nodor bog‘lanishlarni tizimlashtirib, ularni bilimlar bazasi sifatida tashkil etadi. Bu esa avtomatik tarjima, axborot qidiruv va NLP tizimlarining aniqligini sezilarli darajada oshiradi. Tadqiqot natijalari shuni ko‘rsatadiki, ontologik yondashuv korpus lingvistikasini yangi metodologik bosqichga olib chiqadi. O‘zbek tilshunosligida ham bunday tizimlarni rivojlantirish til resurslarini raqamlashtirish va global ilmiy muhitga integratsiya qilish imkonini beradi. Umuman olganda, ontologiya korpus lingvistikasining nazariy asosini mustahkamlovchi va amaliy imkoniyatlarini kengaytiruvchi muhim ilmiy vosita hisoblanadi.

### **Foydalanilgan adabiyotlar ro‘yxati**

1. Abduraxmonova N. *Formal-funksional modellar va o‘zbek elektron korpusi* [Elektron resurs]. – Zenodo, 2021. – URL: <https://doi.org/10.5281/zenodo.5528953>
2. Abjalova M. *Korpus lingvistikasi* [Elektron resurs]. – ResearchGate, 2022. – URL: [https://www.researchgate.net/publication/364348858\\_KORPUS\\_LINGVISTIKASI](https://www.researchgate.net/publication/364348858_KORPUS_LINGVISTIKASI)
3. Abjalova M. *O‘zbek tili ontologiyasi yaratish texnologiyasi va konsepsiyasi* [Elektron resurs]. – ResearchGate, 2022. – URL: [https://www.researchgate.net/publication/363799300\\_OZBEK\\_TILI\\_ONTOLOGIYASI\\_YARATISH\\_TEXNOLOGIYASI](https://www.researchgate.net/publication/363799300_OZBEK_TILI_ONTOLOGIYASI_YARATISH_TEXNOLOGIYASI)
4. Chiarcos C. *An ontology of linguistic annotations* [Elektron resurs]. – Semantic Scholar, 2008. – URL: <https://www.semanticscholar.org/paper/An-ontology-of-linguistic-annotations>



5. Keet C. M. *Ontology Engineering for NLP Systems* [Elektron resurs]. – arXiv, 2017. – URL: <https://arxiv.org/abs/1709.01991>
6. Arabic Ontology Project. *Arabic Ontology based on WordNet* [Elektron resurs]. – arXiv, 2022. – URL: <https://arxiv.org/abs/2205.09664>